BBC RD 1973/36

BBC

RESEARCH DEPARTMENT

1969
THE QUEEN'S AWARD
TO INDUSTRY

REPORT

# A study of high order vertical aperture correction

J.O. Drewery, M.A., Ph.D.

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION

November 1973

**BBC RD 1973/36**
UDC 621.397.6

# A STUDY OF HIGH ORDER VERTICAL APERTURE CORRECTION

**J.O. Drewery, M.A., Ph.D.**

## Summary

*The process of vertical aperture correction involving signal contributions from more than two lines adjacent to the corrected line has been studied theoretically. The conclusion is reached that little is gained by increasing the complexity of correction, and the results of subjective tests confirm this.*

Issued under the authority of

Head of Research Department

**Research Department, Engineering Division,**
**BRITISH BROADCASTING CORPORATION**

November 1973

(PH-118)

# A STUDY OF HIGH ORDER VERTICAL APERTURE CORRECTION

Section Title Page

Summary ........ Title Page

1. Introduction ........ 1

2. Theoretical considerations ........ 1

2.1. The theory of aperture degradation ........ 1
2.2. The vertical aperture correction process ........ 2
2.3. Perfect source correction ........ 4
2.4. Practical source correction ........ 5
  2.4.1. Frequency response ........ 5
  2.4.2. Transient response ........ 7
2.5. The effect of interlace ........ 9
2.6. The effect of source noise ........ 10

3. Practical tests ........ 11

3.1. Introduction ........ 11
3.2. Colour considerations ........ 12
3.3. The measurement of source aperture ........ 14
3.4. Subjective assessments of predicted correction ........ 16
3.5. Subjective assessment of optimally corrected pictures ........ 16

4. Conclusions ........ 18

5. References ........ 18

Appendix 1 ........ 19

Appendix 2 ........ 22

# A STUDY OF HIGH ORDER VERTICAL APERTURE CORRECTION

J.O. Drewery, M.A., Ph.D.

## 1. Introduction

The loss in vertical resolution of television cameras and telecine machines caused by finite spot size is commonly compensated for by using a vertical aperture corrector.[1] This operates by modifying the signal associated with any particular line using information derived from signals on adjacent lines and, until now, the correction information has been derived from the signals on only the two nearest neighbouring lines. The amount of correcting signal can be varied to suit the particular source under consideration but if too much correction is applied, horizontal edges in the displayed picture are accompanied by overshoots and, in extreme cases, by 'ringing' extending for many lines above and below the edges.

The object of the work presented in this report was to investigate the effect of applying more complex correction using information derived from many adjacent lines in the hope that by so doing an improved degree of subjective sharpness could be achieved without the impairment caused by overshoot.

The penalty of increased complexity, apart from cost, is the increase in noise introduced into the corrected signal as more and more components contribute to the final result. However, this is less important than it would seem because the additional noise components are weighted by coefficients which fall off very rapidly as the complexity increases. This will be shown below.

The application of more complex correction requires a means of simultaneous access to the signals on several adjacent lines. This may be achieved by using an arrangement of cascaded line delays. Analogue delays of this length always produce small but measurable signal impairments and such impairments have, so far, limited the complexity of correction. Digital delays, however, provide a means of delaying signals with in principle no degradation and thus of realising correction of any degree of complexity.

The prototype digital line-store standards converter[2,3] offered a means of experimenting with complex aperture correction, as very little modification to the equipment was required to make it operate as an aperture corrector. The modification consisted simply of taking the output directly from the interpolator and adding digital signals corresponding to the synchronising and blanking signals before feeding the signal to the d.a.c. The interpolator coefficient generator had also to be modified slightly so as to give a constant set of values obtaining throughout the whole field.

Such an arrangement gives the experimenter many degrees of freedom and, if undue experimental time is not to be wasted, a basis of theory is required which can guide investigations in the direction of the parameter values most likely to achieve optimum results. Accordingly space is devoted here to a theoretical treatment of aperture correction before practical experiments are described.

## 2. Theoretical considerations

### 2.1. The theory of aperture degradation

A television system scans an object field with a finite aperture and in so doing converts a two-dimensional spatial function into a one-dimensional temporal function or signal. This signal is used at the display to modulate the light passing through a second aperture scanning in sympathy, so recreating approximately the original scene. The theory of scanning involving the use of finite apertures was first deduced by Mertz and Gray[4] and is so central to the particular problem of aperture correction that it will be reviewed here.

The object field can be considered as the superposition of two-dimensional waves having harmonically related spatial frequencies, much in the same way as a one-dimensional function can be broken down into its Fourier components. A general physical two-dimensional spatial frequency has, say, $m$ cycles per picture width and $n$ cycles per picture height and can be expressed as the sum of two complex spatial frequencies having indices $m$, $n$ and $-m$, $-n$. There is however a further physical two-dimensional spatial frequency having $m$ cycles per picture width and $n$ cycles per picture height which has a slope of opposite sign. This can be expressed as the sum of two complex frequencies of indices $m$, $-n$ and $-m$, $n$. A general complex frequency can be denoted simply by the pair $(m, n)$ having an amplitude $A_{mn}$. The scanning action converts each complex spatial frequency $(m, n)$ into a conventional temporal frequency, having an amplitude proportional to $A_{mn}$ and a frequency $f$ given by:

$$f = mf_{\mathrm{H}} + nf_{\mathrm{v}} \tag{1}$$

where $f_{\mathrm{H}}$ and $f_{\mathrm{v}}$ are the horizontal and vertical scanning frequencies respectively. If the picture is horizontally scanned sequentially in $N$ lines $f_{\mathrm{H}} = f_{\mathrm{L}}$, the line frequency, $f_{\mathrm{v}} = f_{\mathrm{p}}$, the picture frequency, and $f_{\mathrm{L}} = Nf_{\mathrm{p}}$. Then Equation (1) becomes

$$f = f_{\mathrm{L}}(m + n/N) \tag{2}$$

If $N$ is very large, Equation (2) shows that the signal spectral energy appears in characteristic bunches centred on multiples of the line frequency as shown in Fig. 1. Equation (2) also shows that although any given spatial frequency $(m, n)$ transforms into one temporal frequency the reverse is not true. For any pair of values $(m', n')$ which satisfy the relationship

$$m' + n'/N = m + n/N$$

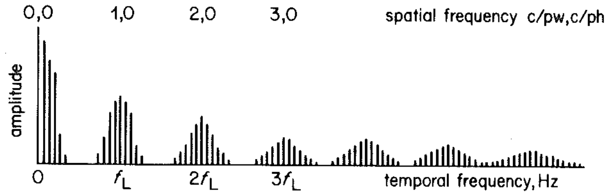


*Fig. 1 - Characteristic form of sequential television signal spectrum*

also corresponds to the same temporal frequency. In other words, if the values of $m$ and $n$ are unrestricted the bunches of energy can overlap. This effect was termed 'confusion' by Mertz and Gray. Nowadays this would be explained by saying that the scanning action spatially samples the scene in the vertical direction so that 'aliasing' can occur if the scene contains components having vertical frequencies greater than half the sampling frequency. Thus the aliasing or confusion can be avoided by restricting the vertical frequencies of the original scene to values of $n$ below $N/2$. This brings out an important distinction in the status of the horizontal and vertical directions, in that it is not necessary to restrict the values of $m$ to avoid confusion. In other words there is no basic sampling in the horizontal direction.

The effect of using a finite scanning aperture is to modify the amplitudes of the spatial frequency components in the scanned signal, related to $A_{mn}$, by a factor $Y(m,n)$ which is the two-dimensional Fourier transform of the reversed scanning aperture.* This is important because it means that it is possible to consider the object field to be pre-distorted by the aperture before scanning. Thus, for example, it is possible to avoid confusion in the signal by choosing a suitable scanning aperture which attenuates spatial frequencies having values of $n$ greater than $N/2$. This is entirely analogous to the concept of appropriately restricting the baseband frequency range to avoid aliasing in a sampling process.

If the original scene contains no components with values of $n$ greater than $N/2$ then clearly the function of aperture correction is to restore the amplitudes of the components in the scanned signal to the values that would have been obtained had the scanning been carried out with a point aperture. If, however, the possibility of confusion exists it is a difficult matter to decide whether or not to restore the components in their entirety and, if not, whether to apply a sharp cut or gentle 'roll-off'.

In principle it is only necessary to pass the signal through a network whose response/frequency characteristic is the reciprocal of $Y(m,n)$, $n<N/2$, expressed in terms of temporal frequency. In practice this is done in two stages known as horizontal and vertical aperture correction. These two processes can, however, provide only an approximation to what is required in certain cases as will now be shown

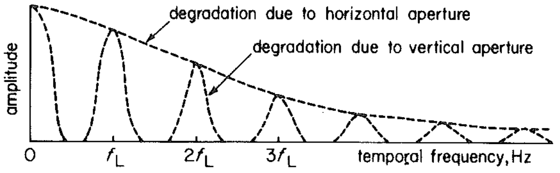


*Fig. 2 - Spectral transfer function of aperture loss*

The general effect of aperture degradation is to attenuate progressively the amplitudes of the higher frequency components. Thus a horizontal aperture corrector has a response which rises steadily throughout the video bandwidth and can be arranged to correct perfectly the components at multiples of line frequency i.e. the horizontal spatial frequencies $(m,0)$. The form of vertical aperture corrector commonly used is the symmetrical transversal filter whose response/frequency characteristic is periodic at multiples of line frequency; thus if the corrector response is unity at zero frequency the spatial frequencies $(m,0)$ are also uncorrected. Roughly speaking therefore the horizontal corrector affects the size or envelope of the spectral energy bunches and the vertical corrector the shape, as shown in Fig. 2. Now implicit in this approach is the assumption that the degradation of the shape of the bunches is uniform throughout the video bandwidth. This in turn implies that the aperture spectrum is variables-separable so that it can be expressed in the form:—

$$Y(m,n) = Y_x(m)\ Y_y(n)$$

where $Y_x(m)$ behaves as an envelope function and $Y_y(n)$ as a shape function independent of $m$. This assumption holds for the rectangular and elliptical Gaussian apertures, but not, for example, for the circular aperture. Thus, the conventional aperture correction process cannot correct perfectly for a circular aperture or any aperture that cannot be expressed in variables-separable form.

In order to correct for such apertures a generalised form of transversal filter would be required whose delay elements would consist of alternating groups of element delays and shortened line delays. Such an arrangement gives independent access to each picture point of a cluster.

In the remainder of this discussion attention will be directed towards the vertical aperture correction process, assuming that it is justifiable to separate the horizontal and vertical effects. Indeed, it is only in this context that the terms horizontal and vertical aperture have any real meaning. The process of vertical aperture distortion and correction can then be treated as a one-dimensional spatial problem involving spatial sampling.

### 2.2. The vertical aperture correction process

The spatial field down a vertical line in the scene can be expressed as a superposition of complex sinusoids of amplitude $A_n$ where the $A_n$'s are functions of the position of the vertical line and the amplitudes $A_{mn}$, that is

---

* If the scanning aperture is $T(x,y)$ the reversed aperture is $T(-x,-y)$.

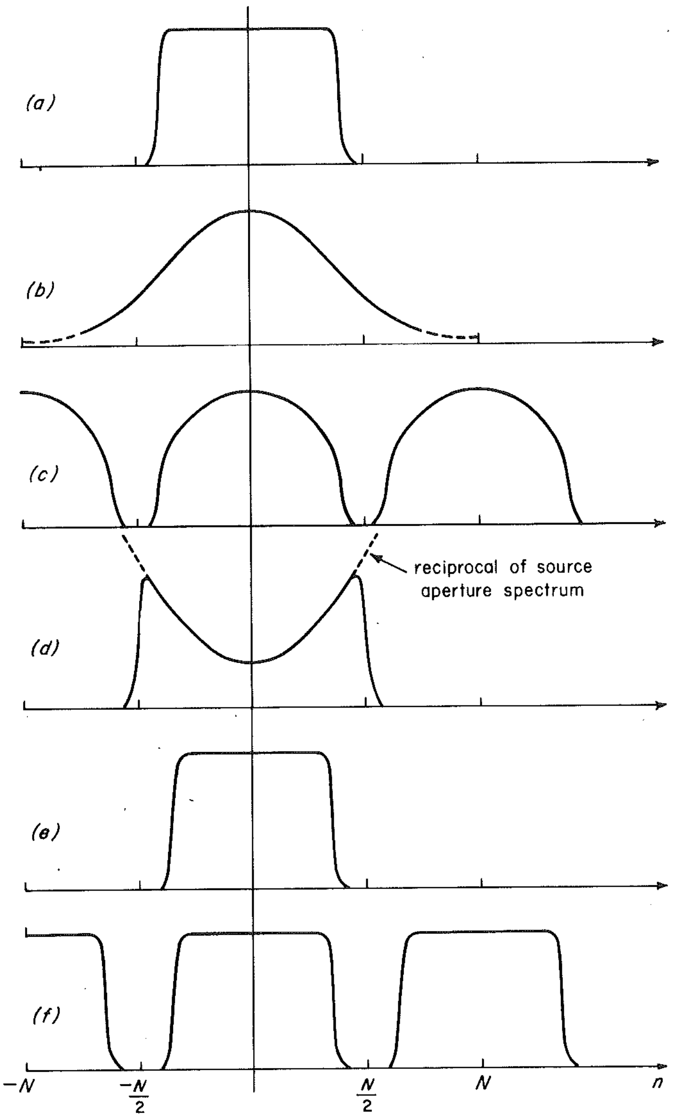


*Fig. 3 - The correction process in the spectral domain*
(*a*) original signal (*b*) source aperture (*c*) sampled signal
(*d*) correction aperture
(*e*) convolution of (*c*) and (*d*) in spatial domain
(*f*) corrected and resampled signal

$$E(x_1,y) = \sum_{n=-\infty}^{\infty} A_n \,.\, \exp j\pi n y/b$$

where
$$A_n = \sum_{m=-\infty}^{\infty} A_{mn} \,.\, \exp j\pi m x_1/a$$

and the scene has dimensions $2a$ and $2b$ along the $x$ and $y$ directions respectively. After scanning with a finite variables-separable aperture the pre-distorted field down the same vertical line could be expressed as

$$E'(x_1,y) = \sum_{n=-\infty}^{\infty} A_n' \,.\, Y_y(n) \,.\, \exp j\pi n y/b$$

where
$$A_n' = \sum_{m=-\infty}^{\infty} A_{mn} \,.\, Y_x(m) \,.\, \exp j\pi m x_1/a$$

Assuming that the horizontal degradation is subsequently corrected so that $A_n'$ becomes $A_n$, the effect of vertical degradation is to multiply the vertical spectrum by the function $Y_y(n)$, the spatial Fourier transform of the vertical aperture. Thus it is possible to frame the discussion in terms of the one-dimensional spectral functions $A_n$ and $Y_y(n)$.

In order to correct the degradation it is necessary to operate on the signal, which gives values of the degraded image at equispaced points down the vertical line, by restoring these values to those that would have obtained had the scanning been carried out with a point aperture. Remembering that sampling action results in a spectrum with sidebands centred on multiples of the sampling frequency, the sequence of operations involved in correction could be described in the spectral domain as in Fig. 3. The original signal whose vertical spectrum, $A_n$, is shown at (*a*) is scanned with an imperfect source aperture whose vertical spectrum, $Y(n)$, is shown at (*b*) to produce a signal whose spectrum is shown at (*c*). Multiplication by the function shown at (*d*) produces the spectrum shown at (*e*) which can then be rescanned in the space domain to produce a signal approximating to perfect scanning whose spectrum is shown at (*f*).

It will be observed that the correction function shown at (*d*) departs from the reciprocal of the source aperture spectrum at frequencies above about half the sampling frequency. This must be so, for two reasons; first, as the source aperture spectrum falls to zero the reciprocal cannot tend to infinity and, secondly, the higher spectral orders must be eliminated if they are not to intrude into the zero-order spectrum after resampling (i.e. by aliasing).

The correction process represented by the step from (*c*) to (*f*) has a transfer function which is periodic in the spatial frequency domain and is, in fact, obtained by superposing versions of the function shown at (*d*) repeated at intervals of $N$. Now the symmetrical transversal filter shown in Fig. 4(*a*) can, with certain values of the coefficients, $c_r$, be given the response shown in Fig. 4(*b*). If the value of $\tau$ is a line period the transformation of spatial to temporal frequency implied by Equation (2) means that there is clearly a relationship between the function shown at Fig. 3(*d*) and the coefficients $c_r$, as will now be shown.

The multiplication of the spectrum of Fig. 3(*c*) representing the degraded samples by the function shown at (*d*) is equivalent to convolving the samples with a function which is the inverse Fourier transform of that shown at (*d*). That is to say, each sample is replaced by a scaled version of this inverse Fourier transform which, by analogy, could be called a correction aperture. The continuous function so produced is then resampled at the same sample points to yield modified samples. This process is shown in Fig. 5 which is the spatial equivalent of Fig. 3. As, in general, the correction aperture will not have zeros at the sampling

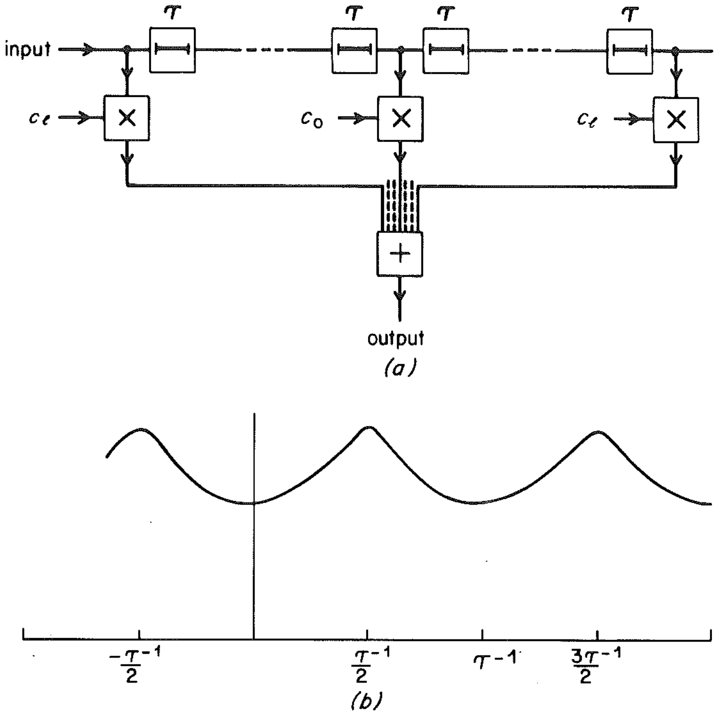


*Fig. 4 - The transversal filter*
(*a*) network (*b*) a possible transfer characteristic

points, the magnitude of each sample will modify its neighbours. Now returning to the transversal filter of Fig. 4(*a*), if the input signal is such that the signal magnitude is zero at all the tapping points except one, as the signal advances steadily through the filter the output of the filter is that signal magnitude multiplied by successive coefficient values. But this is exactly what would have been obtained by convolving the isolated sample with the correction aperture and resampling the result at the sample points. Thus the coefficient values are the values of the correction aperture at the sample points.

The corrector is adjusted by varying the values of the coefficients until the response over the zero-order spectral region approximates to $[Y_y(n)]^{-1}$. It will now be realised that in order to correct for an aperture perfectly up to the theoretical limit of displayed spatial frequency, that is $N/2$ cycles/picture height, the corrector response must have a cusp at this frequency because the response is periodic. This in turn, implies that there must be an infinite number of coefficients. Looked at another way, for perfect correction the function shown at Fig. 3(*d*) must fall discontinuously to zero at the limit. Thus the correction aperture, being the inverse Fourier transform, will 'ring' indefinitely producing an infinite number of coefficients. Thus, not unexpectedly, a perfect aperture correction requires a corrector of infinite extent.

Although the goal of perfect correction can never be achieved, it is nevertheless instructive to calculate the ideal coefficient values for an infinite corrector because they indicate the order of magnitude of the quantities involved.

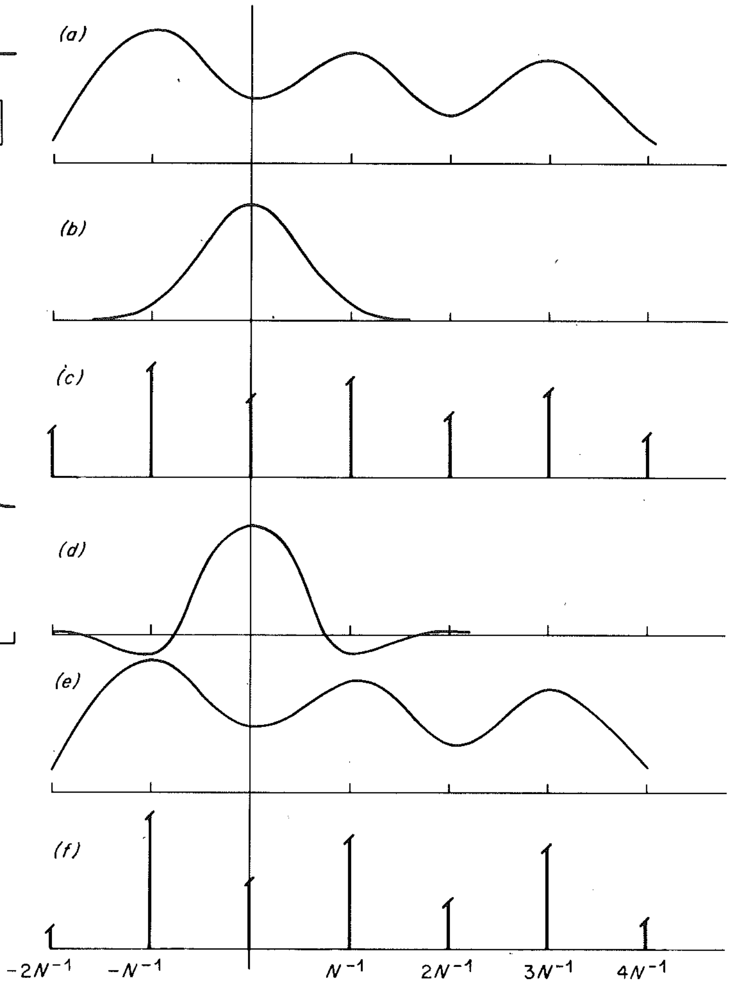


*Fig. 5 - The correction process in the spatial domain*
(*a*) original signal (*b*) source aperture (*c*) sampled signal
(*d*) correction aperture (*e*) corrected signal
(*f*) corrected and resampled signal

### 2.3. Perfect source correction

The preceding section indicates a method of deducing coefficient values. The steps are:—

Find the spatial Fourier transform of the vertical aperture.

Take the reciprocal and truncate it so that it extends over a region of width equal to the spatial scanning frequency.

Take the inverse Fourier transform.

Calculate values of the function at the sampling (scanning) points.

This method can be applied to any vertical aperture, provided that its spatial Fourier transform does not have a zero within the range $\pm N/2$ c/p.h. Fig. 6 shows the values of the first few coefficients when the method is applied to rectangular and Gaussian apertures. The derivation of these curves is given in Appendix 1 where it will be seen

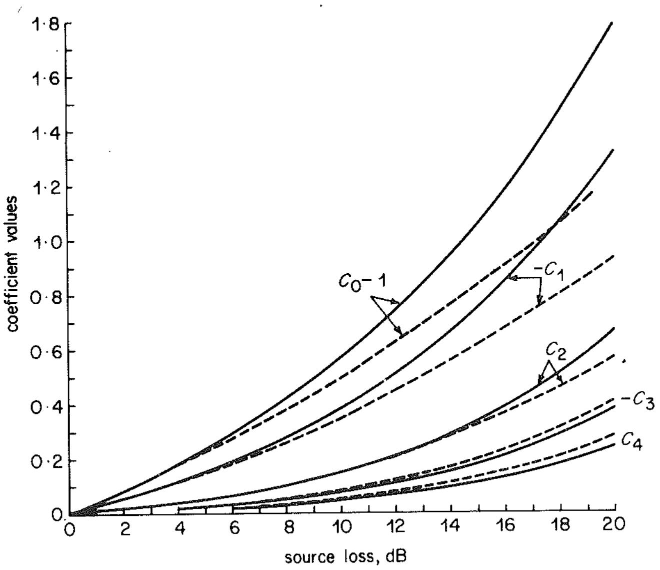


*Fig. 6 - Theoretical coefficient values for perfect source aperture correction*

——— Gaussian - - - - - rectangular

that the taking of the inverse Fourier transform is done by expanding the correction aperture spectrum as an even power series within the range $\pm N/2$. To a good approximation, however, it is only necessary to take the first two terms of the series if the amount of correction is low, and expressions for the coefficients are obtained as follows:

$$c_0 = 1 + (\pi w)^2/72$$

$$c_r = (-1)^r w^2/12r^2$$

for the rectangular aperture of width $w/N$ p.h.

and
$$c_0 = 1 + (\pi s)^2/6$$

$$c_r = (-1)^r s^2/r^2$$

for the Gaussian aperture of standard deviation $s/N$ p.h.

The values $w$ and $s$ can be related to $D$, the dB of correction needed at half sampling frequency, $N/2$, by the equations

$$D = 20 \log_{10} (1 + \pi^2 w^2/24)$$

or
$$= 20 \log_{10} (1 + \pi^2 s^2/2)$$

Note that the simple approximation causes the coefficients to fall off as the inverse square, in both cases.

This result is somewhat surprising as the 'ringing' of a sin$x/x$ pulse dies away as the inverse.

### 2.4. Practical source correction

#### 2.4.1. Frequency response

A practical corrector has an aperture of finite width which means that the function of Fig. 3(*d*), being the Fourier transform of the correction aperture, is unbounded. It is therefore more helpful to derive the correction coefficients by way of the transversal-filter response, rather than the correction aperture and its transform, because of the overlapping of the correction-aperture spectra. Now it can be shown that the spectral amplitude response of the transversal filter is given by the expression

$$H(\theta) = c_0 + 2 \sum_{r=1}^{l} c_r \cos r\,\theta$$

where $\theta$ is an angular co-ordinate that varies from $-\pi$ to $+\pi$ across the zero-order spectrum, that is for spatial frequencies in the range $\pm N/2$ c/p.h. The relationship between $\theta$ and $n$ is thus given by

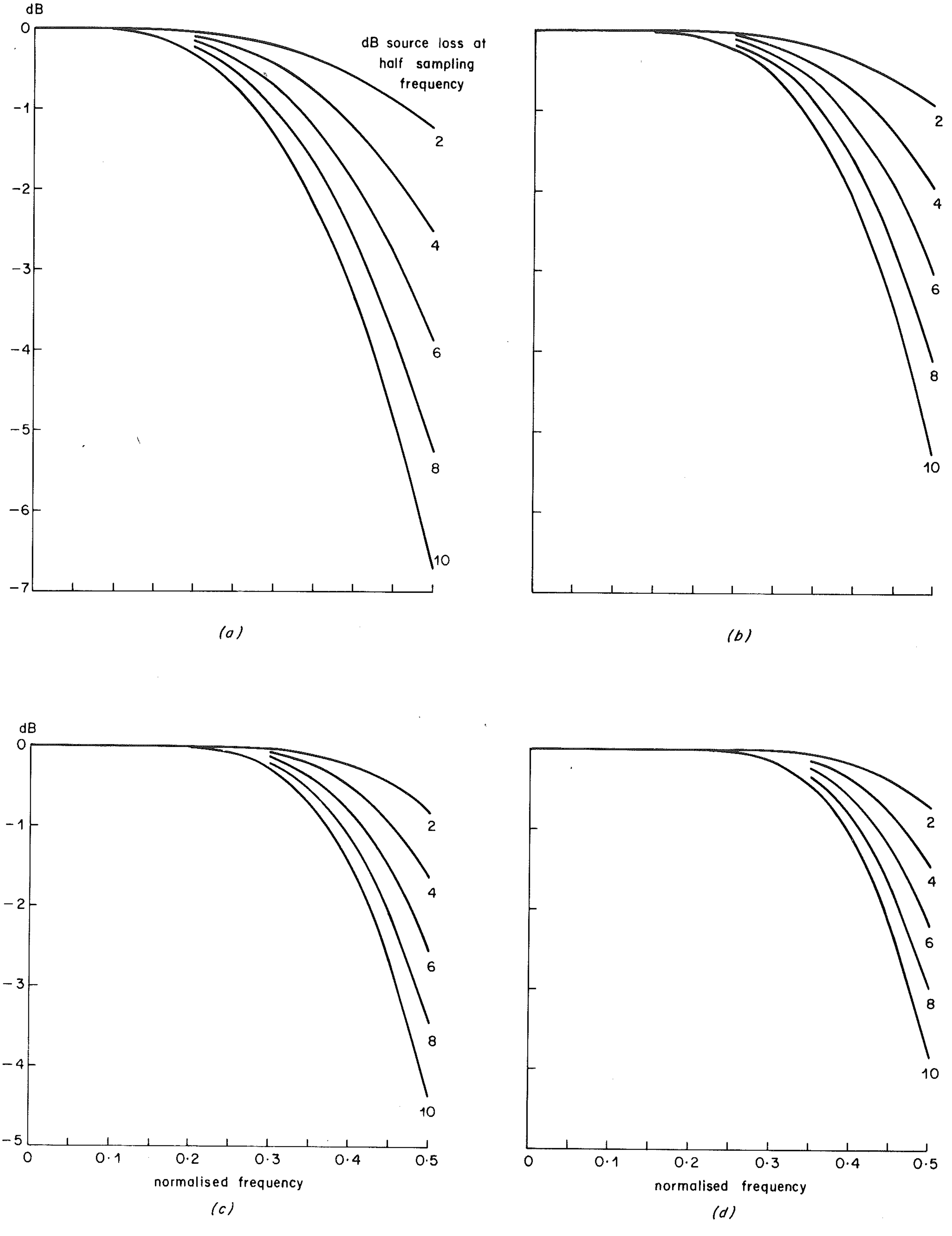
$$\theta = 2\pi n/N$$

If the signal-component amplitudes are restored to their correct values the spectral response of the corrector is the reciprocal of the scanning-aperture spectrum over the zero-order spectral region, that is

$$Y_y(n) H(\theta) = 1, \; |\theta| < \pi$$

However, as previously noted, unless the corrector response has a cusp at the point $\theta = \pi$ and therefore an infinite number of terms, it cannot correct perfectly. Thus the function $Y_y(n)H(\theta)$ cannot equal unity over the whole region and the error can be expressed as a deviation from unity. The difficulty now arises as to where to allocate the error. For example a positive error in one part of the spectrum can be offset by a negative error in another part without causing undue subjective impairment. The simplest yardstick by which to judge the efficacy of correction as a function of $l$, the order of correction, is to allocate the error to the higher frequencies by making the overall response maximally-flat at zero frequency; this will be referred to as the maximally-flat condition. It can be achieved by expanding both $[(Y_y(n)]^{-1}$ and $H(\theta)$ as even power series and equating coefficients of ascending powers. As both power series are infinite this operation produces an infinite number of equations containing $l + 1$ unknowns and for the maximally-flat condition the first $l + 1$ equations are satisfied. An example of this process for the Gaussian source aperture is given in Appendix 2. When the coefficient values are found they can be substituted in the function $H(\theta)$ and the overall response $Y_y(n)H(\theta)$ can be determined.

Fig. 7 shows the overall response/frequency characteristics obtained for values of $l$ up to 4, that is, 4th-order correction. A Gaussian source aperture is assumed with various values of loss at the half-sampling frequency. As can be seen the characteristics never overshoot and the effect of adding higher-order correction is to delay the onset of the roll-off. There would appear to be a law of diminishing returns in that the improvement gained by applying first-order correction is more than twice that gained by adding second-order correction. Fig. 8 shows the corrector response alone for values of $l$ up to 4 and for 10



(a) (b) (c) (d)

*Fig. 7 - Overall response of source aperture and corrector. Gaussian source with corrector adjusted for maximally-flat response at zero frequency*

Various orders of correction (*a*) 1st-order (*b*) 2nd-order (*c*) 3rd-order (*d*) 4th-order

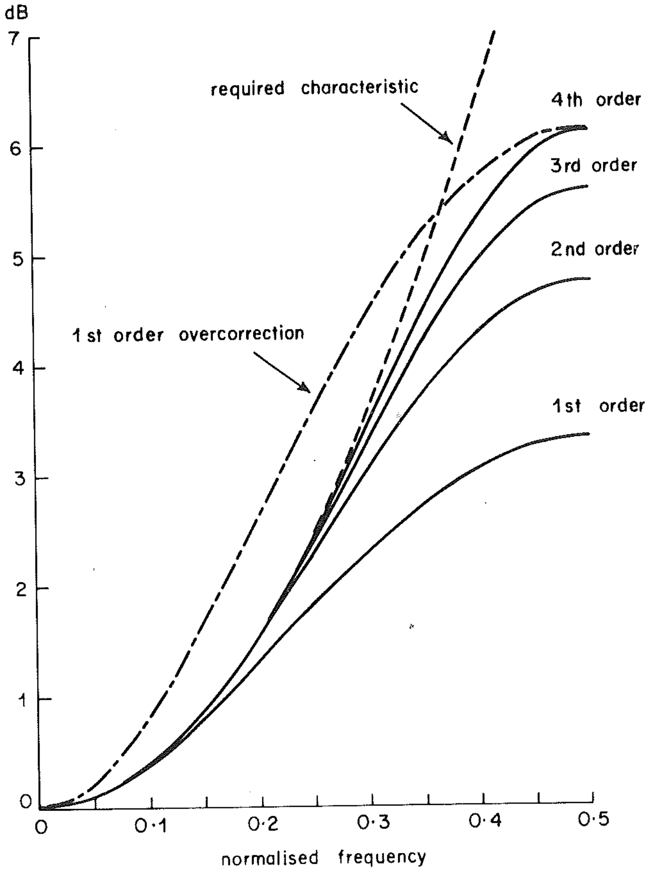


*Fig. 8 - Corrector response adjusted for maximally-flat correction of Gaussian source at zero frequency. Source loss of 10 dB at half-sampling frequency*

dB of source loss at the half-sampling frequency. Note the increasing sharpness of response at the half-sampling frequency as $l$ increases.

### 2.4.2. Transient response

Although the response/frequency characteristic is a measure of definition it gives no indication of the transient behaviour of the system and in particular the degree of overshoot which results from correction. The transient behaviour can be deduced starting with an appropriate signal spectrum and working through the system operations in the spectral domain, finally transforming back to the space/time domain. This process involves the display aperture mentioned earlier and the form of the display aperture considerably influences the final result, as will be seen.

Fig. 9 shows a computer simulation of the system response to a $2T$ edge; such an edge is defined as the result of passing an infinitely sharp edge through a filter having a $2T$-pulse response where $T$ is the period of the highest frequency that can be displayed by the system. In vertical spatial terms $1/T$ is equivalent to $N/2$ and $1/N$ corresponds to a vertical picture element (PEL). Thus the $2T$ edge under consideration has a width of $2T$ or 4 PELs. Such an edge has been chosen because it contains very little energy above the spatial frequency $N/2$. A source loss of 10 dB at the half-sampling frequency and a display aperture having a flat response/frequency characteristic cutting sharply at the half-sampling frequency have been assumed. The use of this display aperture, although physically unrealisable as it is of $(\sin x)/x$ form which requires negative luminance, is nevertheless helpful in indicating the general behaviour of the system in response to different degrees of correction without introducing unnecessary detail. The correction coefficients have been derived according to the criterion explained above giving a maximally-flat response at zero frequency.

Because the signal spectral energy is largely confined below half the sampling frequency there is virtually no difference between the original signal and a perfectly sampled and displayed version of it as can be seen from Fig. 9. If, however, a source loss of 10 dB is introduced and remains uncorrected the response is appreciably softer. Using the '10–90%' criterion, the rise-time is increased by some 20%. As can be seen the introduction of first-order correction practically restores the rise-time to its former value with a 0·6% overshoot. After 4th-order correction the rise-time is indistinguishable from the original but an overshoot of 0·3% remains. The remaining curve indicates the response when the first-order correction is increased to the point where the correction at the half-sampling frequency is equal to that of 4th-order correction (based on the criterion of maximally-flat response). This condition is indicated in Fig. 8. Thus the over-corrected first-order and 4th-order responses could perhaps be said to have the same definition for high frequencies. But clearly the over-corrected first-order response gives much more overshoot than the 4th-order. This exercise illustrates the importance of allocating the response error to the higher frequencies.

Figs. 10 and 11 show the system responses to half-sinusoidal edges of widths 3 and 2 PELs respectively. The spectrum of the first edge has a zero at the half-sampling frequency. The same conditions as in Fig. 9 have been assumed. Here, even with perfect sampling and display, the response rings because the signal contains significant spectral energy beyond the half-sampling frequency. The effect is partly caused by aliasing in the sampling operation but predominantly by the sharp cut in the assumed display response. In Fig. 10, as before, first-order correction makes a significant improvement which is little bettered by 4th-order correction. For the edge in Fig. 11, which clearly defeats the system, first-order correction would probably be preferred to 4th-order correction because it reduces the overshoots although improving rise-time considerably.

Lest the foregoing examples seem suspect because of the physically unrealisable display aperture, Fig. 12 shows the same conditions as Fig. 11 but with a Gaussian display aperture having a loss of 10 dB at the half-sampling frequency. The same general conclusions apply although the issue is confused by the presence of the perturbation at the sampling frequency. This perturbation is present because the display aperture attenuates the sampling frequency by only 40 dB instead of completely rejecting it.

In summary a maximally-flat response at zero frequency would seem to result in least overshoot on edges

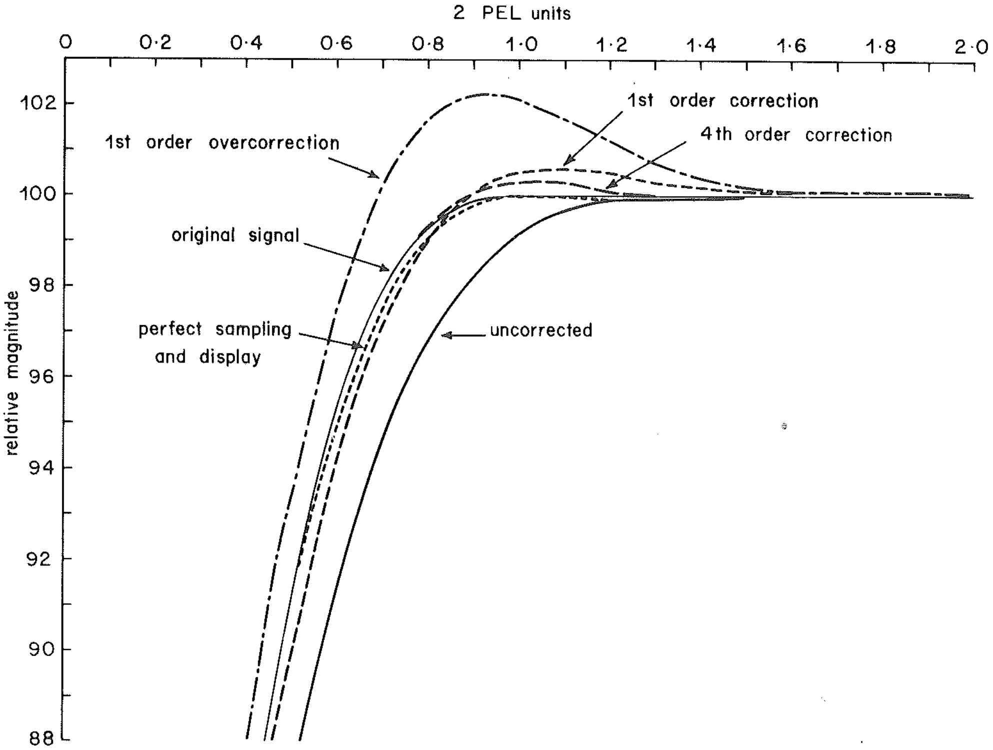


*Fig. 9 - System response to 2T edge with various orders of correction. Gaussian source with loss of 10 dB, sharp-cut display*

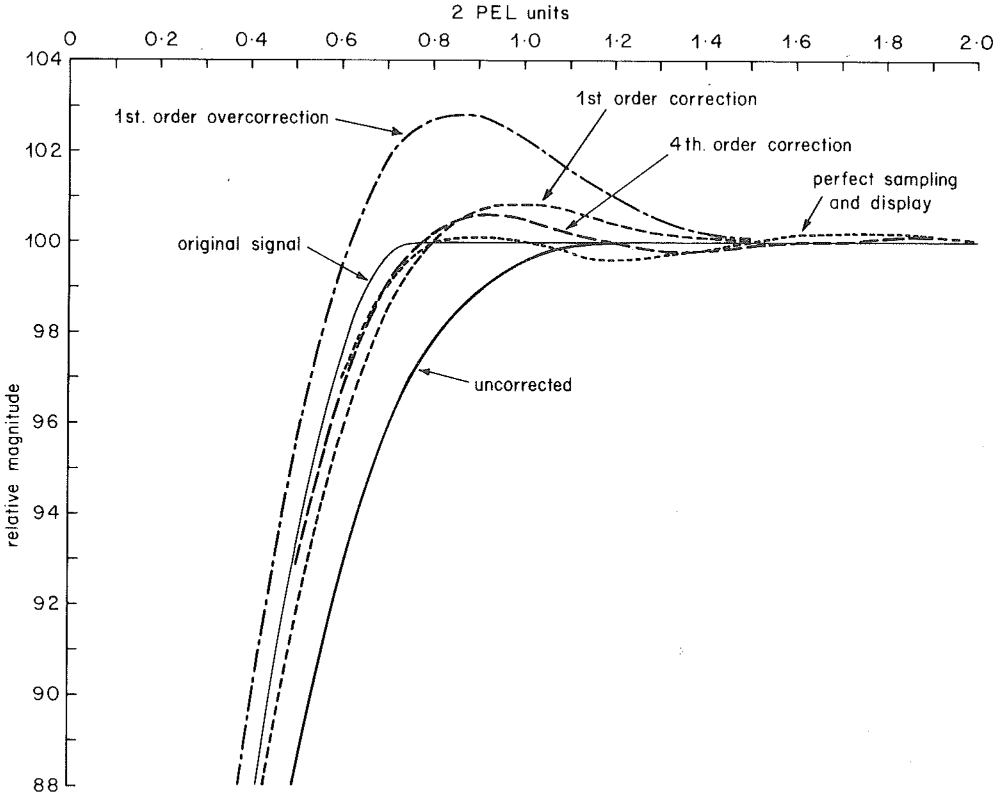


*Fig. 10 - System response to half-sinusoidal edge of width 3 PELs. Other conditions as in Fig. 9*

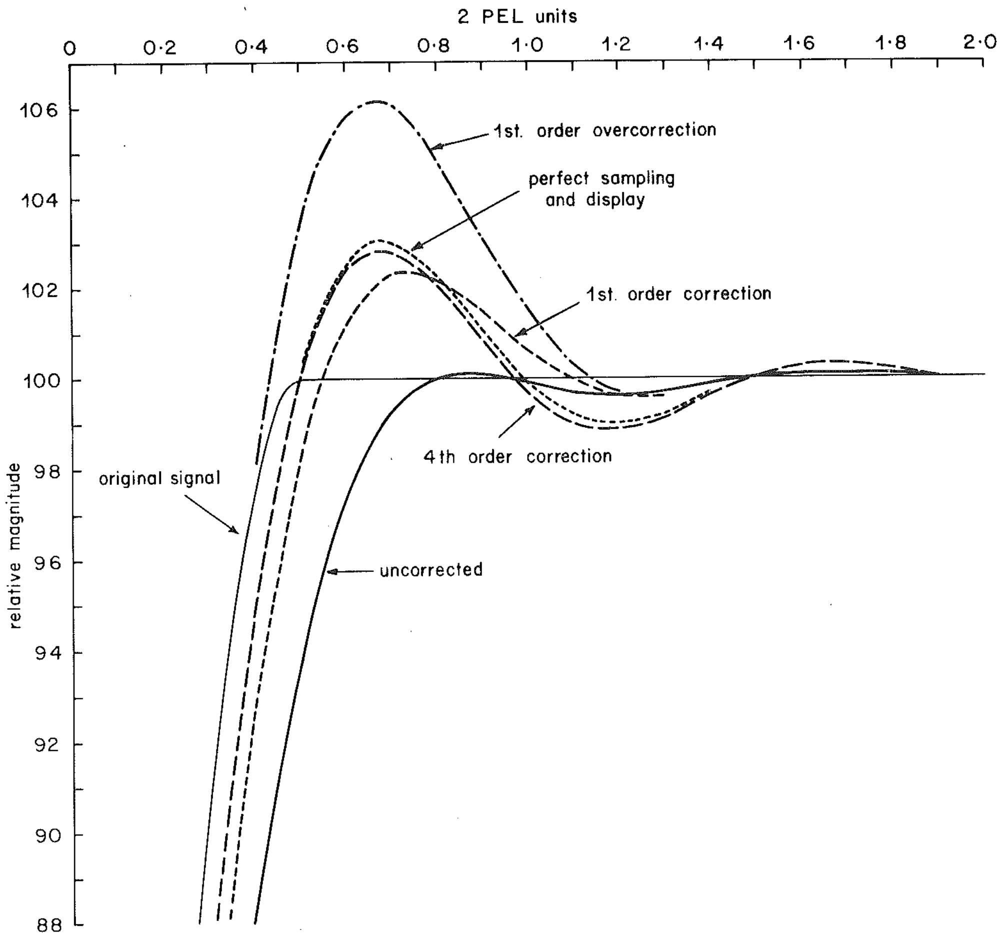


*Fig. 11 - System response to half-sinusoidal edge of width 2 PELs. Other conditions as in Fig. 9*

that are capable of being displayed by the system within the confines of the responses obtainable by finite-order correction. Little is gained by increasing correction beyond first-order. This degree of correction is also probably best for edges which are too sharp for the system and which consequently ring when perfectly corrected.

### 2.5. The effect of interlace

The previous discussion has been concerned with a system involving sequential sampling. The argument need only be slightly modified, however, to deal with interlaced sampling. For simple interlaced sampling the spatial sampling frequency in each field is halved but the odd spectral orders of the interlaced sampled field are inverted.$^5$ Consequently when the spatial spectra of the two displayed fields are combined the odd orders partially cancel to an extent which depends on the temporal response of the display and the observer. Thus the odd orders are attenuated and alternate during a complete picture and, taken over the integration time of a single picture, they can be ignored. The effective spatial sampling frequency over a complete picture is therefore equal to that of sequential sampling. If, however, aperture correction is performed using a transversal filter having access to the samples of only one field the spatial sampling frequency appropriate to the corrector response corresponds to that of the field. Thus the periodicity of the response in the spatial frequency domain is halved and is given by

$$H(\theta) = c_0 + 2\sum_{r=1}^{l} c_r \cos 2r\theta \qquad , \theta = 2\pi n/N$$

which has a period of $N/2$ where, as before, a picture (2 fields) contains $N$ lines. As

$$H(0) = c_0 + 2\sum_{r=1}^{l} c_r = 1$$

for unity gain at zero frequency it is clear that the corrector offers no correction at the frequency $N/2$ c/p.h., the effective half-sampling frequency for the picture. In fact the corrector exerts its maximum effect at the frequency $N/4$ c/p.h., the mid-band of the effective display response. If therefore the device attempts to correct a Gaussian source aperture up to this frequency on a maximally-flat basis the simultaneous equations which govern the correction coefficients are obtained, as before, by equating ascending

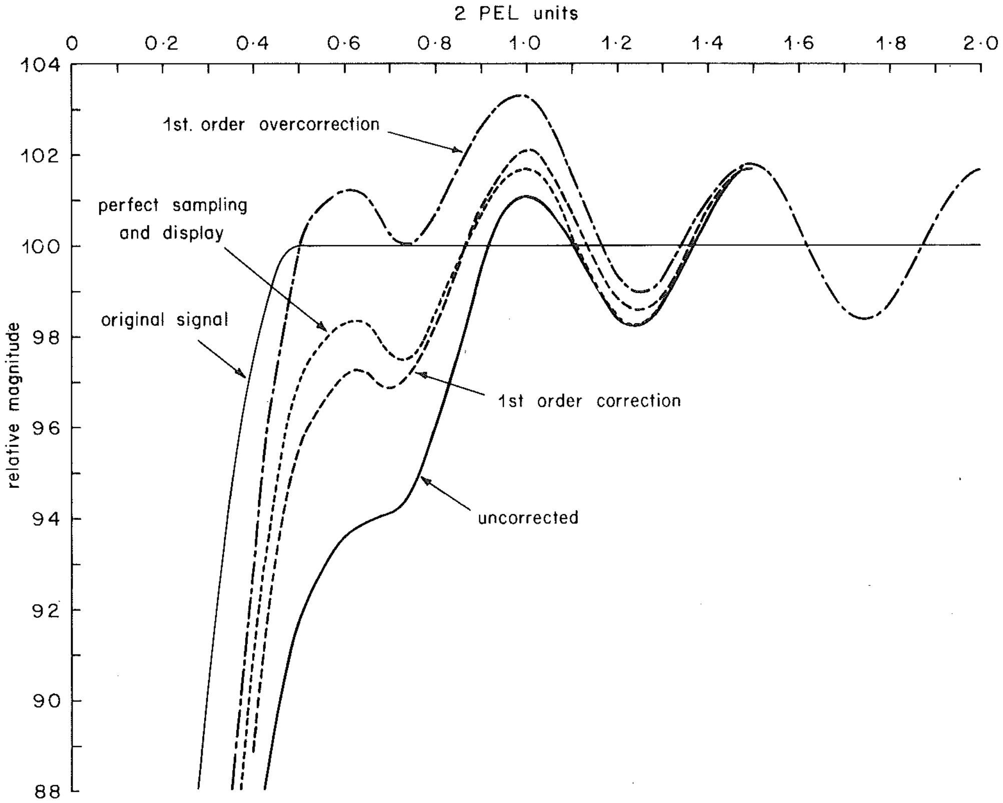


*Fig. 12 - As in Fig. 11 except Gaussian display aperture with 40 dB rejection of sampling frequency*

powers but halving the value of $s$, the normalised standard deviation.

If, as before, the overall response is maximally-flat at zero frequency the first $l + 1$ equations are satisfied. Fig. 13 shows the overall response/frequency characteristics for correction up to 4th-order, and for 10 dB source loss at the half-sampling frequency. Note that there is now no correction at this frequency. Fig. 14 shows the corrector response alone.

Figs. 15, 16 and 17 show the time domain responses to $2T$ and half-sinusoidal edges corresponding to Figs. 9, 10 and 11. The same conditions are assumed but the corrector response is now that of Fig. 14. It will be observed in Fig. 15 that the first-order corrected signal overshoots a little more (0·8%) than in Fig. 9. The 4th-order corrected signal overshoots even more however and is accompanied by further ringing extending away from the edge. This condition would be likely to be more objectionable. The first-order overcorrection is not as extreme as in Fig. 9 as might be expected because of the inability of the corrector to affect the higher frequencies. For the sharper edges of Figs. 16 and 17 the difference between perfect and 4th-order correction becomes more marked than in Figs. 10 and 11 whereas the difference between first and 4th-order correction decreases. The overshoots are also worse than for the sequential case.

In summary the law of diminishing returns seems to apply more strongly for correction involving interlaced sampling than for that involving sequential sampling.

### 2.6. The effect of source noise

As the complexity of the correction increases, more and more signals contribute to the output of the corrector. These signals inevitably contain noise and assuming the noise components are uncorrelated[6] their powers rather than amplitudes add together. Thus the signal-to-noise ratio at the output of the corrector is less than that at the input and the degradation can be calculated as follows.

If the r.m.s. noise voltage at the input is $e_i$ the output r.m.s. noise voltage, $e_o$, is given by

$$e_o^2 = c_0^2 e_i^2 + 2c_1^2 e_i^2 + \ldots\ldots$$

$$= e_i^2\left(c_0^2 + 2\sum_{r=1}^{l} c_r^2\right)$$

This noise will be most visible on areas of uniform luminance corresponding to a constant signal input. Under these conditions the corrector gain is unity and the ratio of the input to output signal-to-noise ratios is simply the ratio of the noise powers. Thus the corrector noise factor in dB is given by

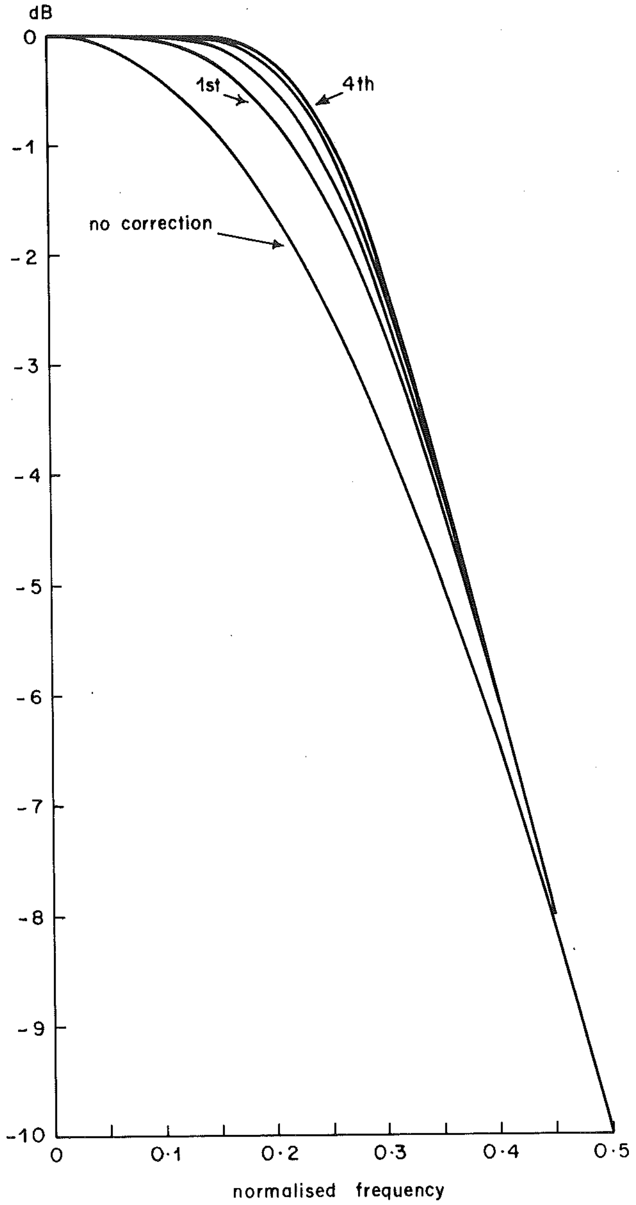


*Fig. 13 - Interlace corrector. Overall response of source aperture and corrector. Gaussian source with loss of 10 dB at half picture sampling frequency. Corrector adjusted for maximally-flat response at zero frequency. Various orders of correction*

$$F = 10\log_{10}\left(c_0{}^2 + 2\sum_{r=1}^{l} c_r{}^2\right)$$

This relationship is plotted in Fig. 18 for various orders of correction and varying degrees of correction. As a corrector can be given any noise factor by appropriate choice of coefficients the criterion of comparison in Fig. 18 is the maximally-flat condition assuming a Gaussian source whereby the coefficient values are determined by the degree of correction needed at the half-sampling frequency. Although the coefficient values fall off rapidly as $r$ increases, a higher-order correction allows the values of all coefficients to rise for the same degree of correction. Thus for 10 dB of correction the value of the main term $c_0$ rises from 1·23 for a first-order correction to 1·56 for perfect correction. As a result the addition of higher order terms does noticeably affect the noise factor of the corrector. However, the increase in noise factor between third and fourth-order correction is much less than that between first and second; but fourth-order correction is still much less noisy than perfect correction. If a figure of merit is defined as the difference in dB between the maximum correction offered (as deduced from Fig. 8) and the noise factor for the same degree of correction, this quantity gives a more meaningful comparison between orders. In fact the quantity increases with increasing order of correction but once again a law of diminishing returns applies.

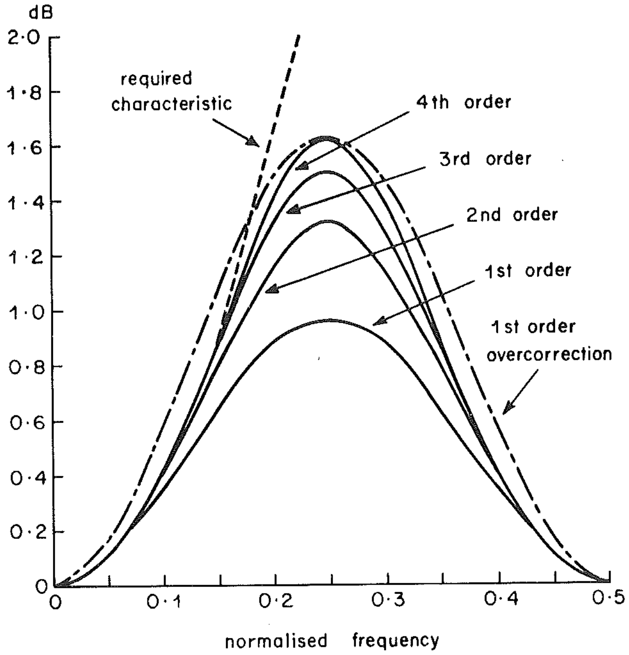


*Fig. 14 - As Fig. 13 - Corrector response alone*

In summary, as the correction becomes more complex, the increase in noise follows roughly the same pattern as the increase in correction, so that, other things being equal, there is little to choose between various complexities. If, however, the noise at the input were always imperceptible there could be a clear benefit in introducing more complex correction.

## 3. Practical tests

### 3.1. Introduction

The foregoing theoretical analysis implies that if the source aperture is corrected so that the spectral response is maximally-flat at zero frequency there is little to be gained by introducing higher-order terms to a first-order correction process. It may be, however, that the maximally-flat condition may not be, subjectively, the best form of correction. Accordingly two methods of approach suggest themselves. (A) The necessary amount of correction for different orders can be predicted according to the maximally-flat condition

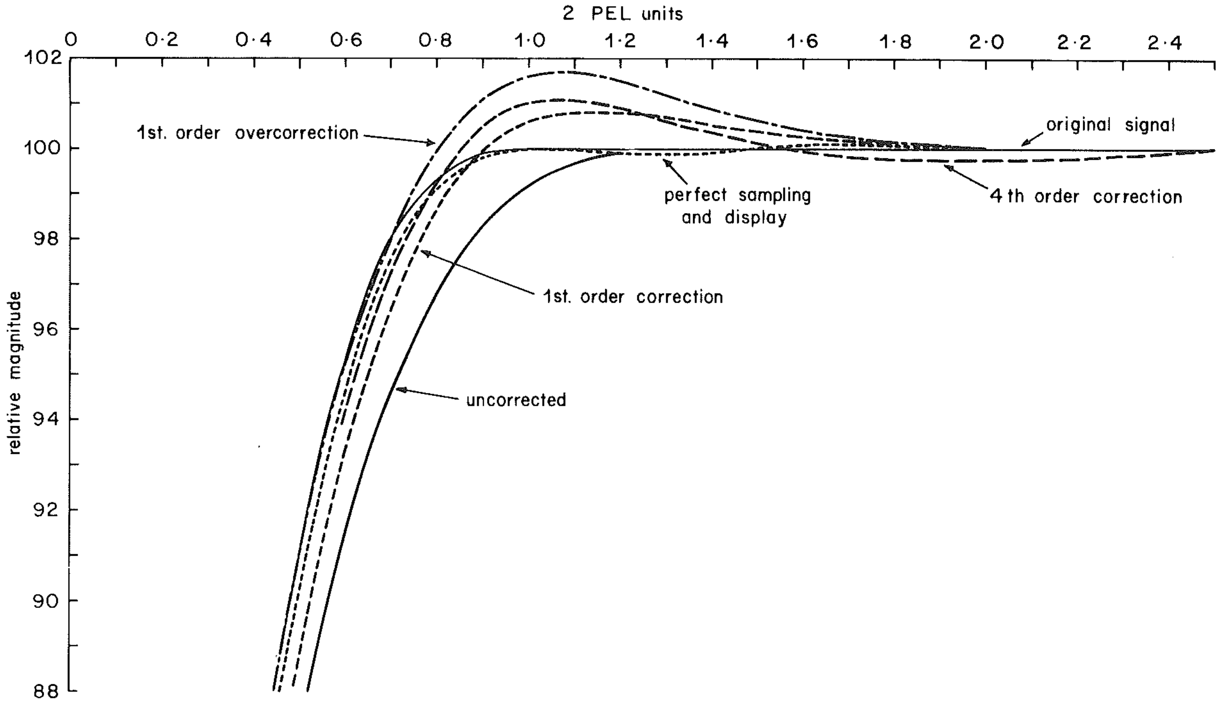


*Fig. 15 - Interlaced system response to 2T edge. Conditions as in Fig. 9*

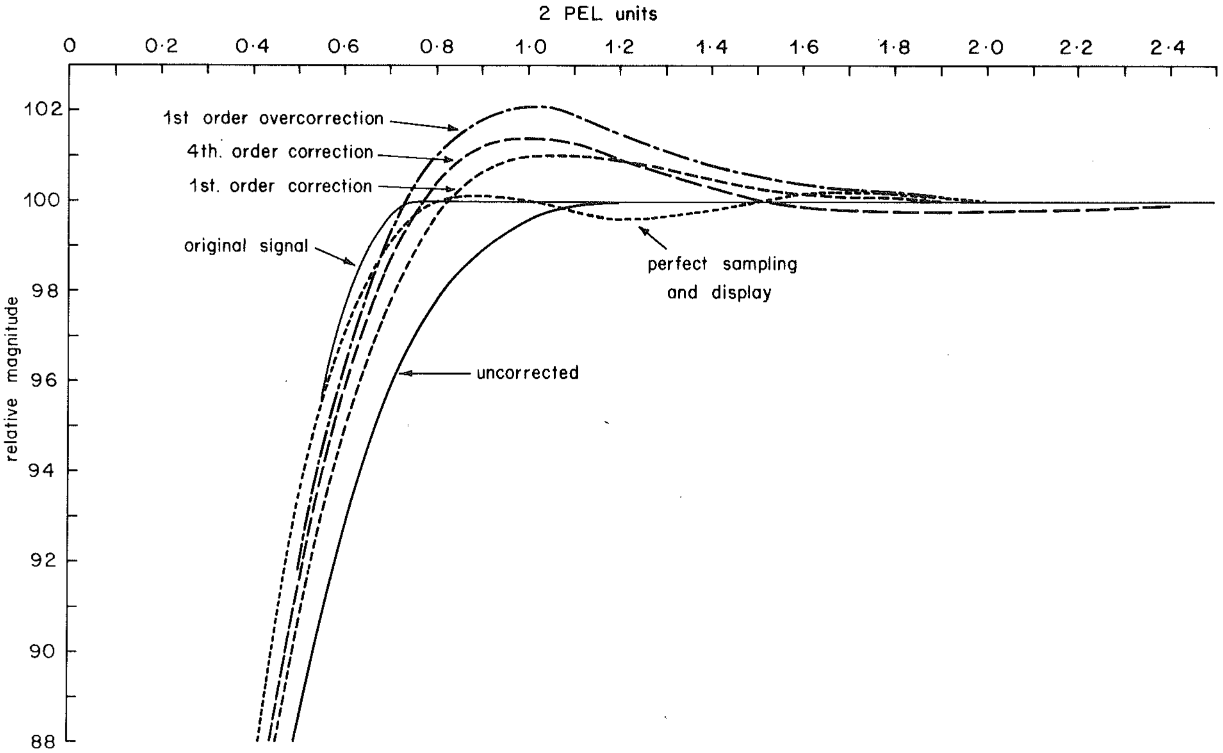


*Fig. 16 - Interlaced system response to half-sinusoidal edge of width 3 PELs. Conditions as in Fig. 9*

knowing the source characteristics, and different orders can be compared subjectively. (B) The subjectively optimum conditions for each order of correction can be compared. Both these methods were tried in an attempt to investigate the benefit of adding higher-order correction.

### 3.2. Colour considerations

A preliminary investigation was carried out to test whether the results of correcting monochrome pictures could be applied to the correction of colour pictures. A

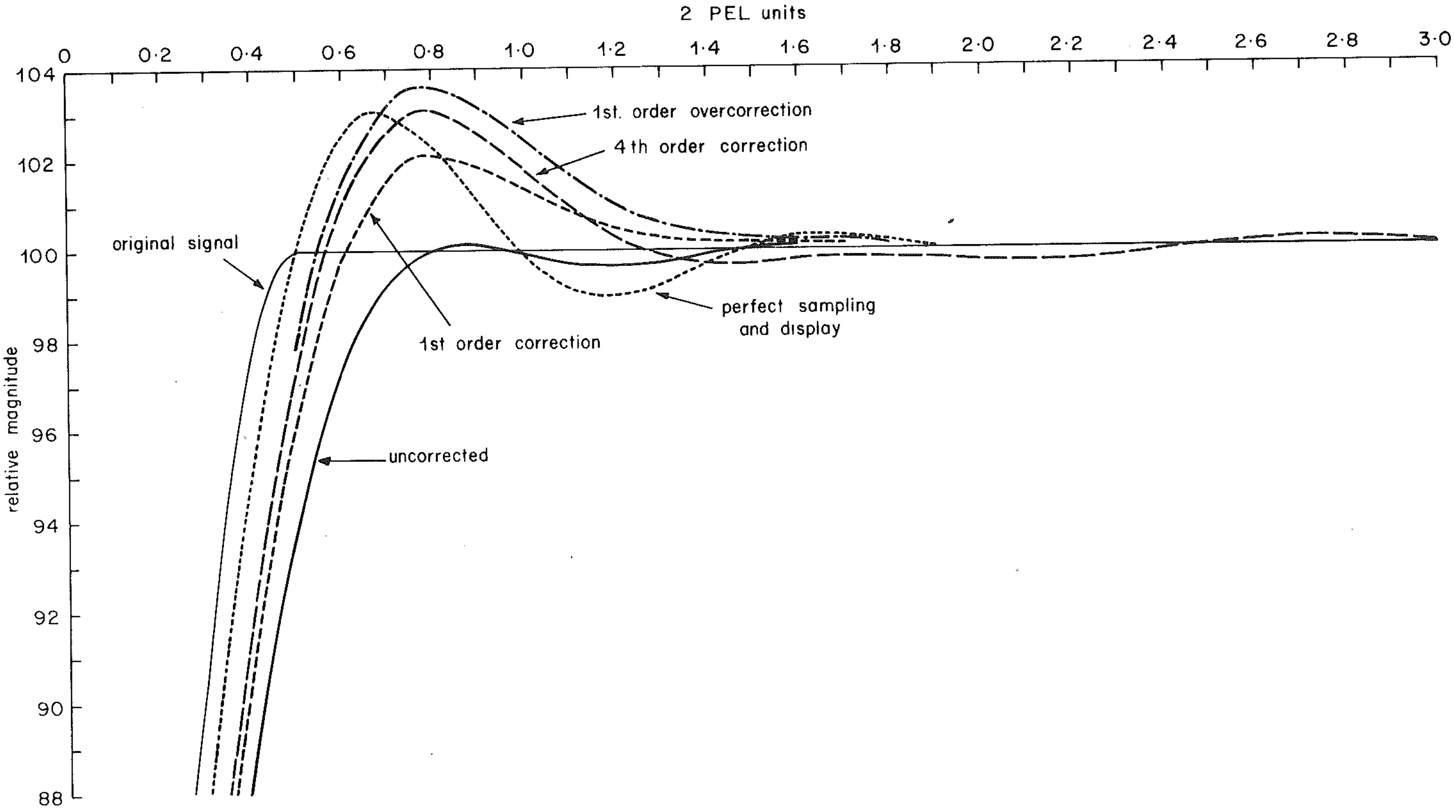


*Fig. 17 - Interlaced system response to half-sinusoidal edge of width 2 PELs. Conditions as in Fig. 9*

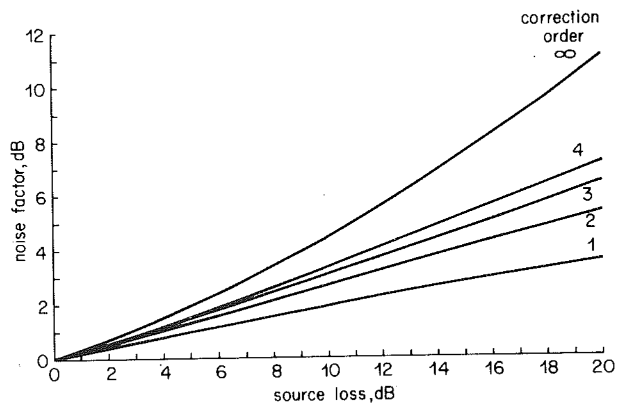


*Fig. 18 - Corrector noise factor as a function of source loss for various orders of correction*

4-tube plumbicon camera was used as the source and a digital vertical-aperture corrector was inserted in the feed of linear luminance signal to the processing circuits. The luminance camera tube was misregistered by an amount corresponding to the delay introduced by the corrector so that the corrected luminance signal was registered with the signals from the other three tubes. Two picture subjects were used. The first was a bunch of highly coloured flowers containing much fine detail. The second was an isolated flower with a thin, near-horizontal leaf against a light background. Five different states of correction were used, shown in Table 1, and six observers were asked to score the pictures, for definition and overshoot separately, according to the EBU impairment scale shown in Table 2. The colour was then removed from the picture (by shorting together the display guns to avoid cross-colour) and the observers were asked to score the same states of correction.

*TABLE 1*

*Correction Characteristics Used in Colour Tests*

| Correction Condition | Coefficients | | | dB loss at half-sampling frequency | |
|---|---|---|---|---|---|
| | $c_0$ | $c_1$ | $c_2$ | Field | Picture |
| 0 | 1·00 | 0 | 0 | 0 | 0 |
| 1 | 1·23 | −·155 | +·04 | 8 | 32 |
| 2 | 1·34 | −·17 | 0 | 16 | 64 |
| 3 | 1·90 | −·60 | +·15 | 24 | 96 |
| 4 | 2·70 | −·85 | 0 | >40 | >160 |

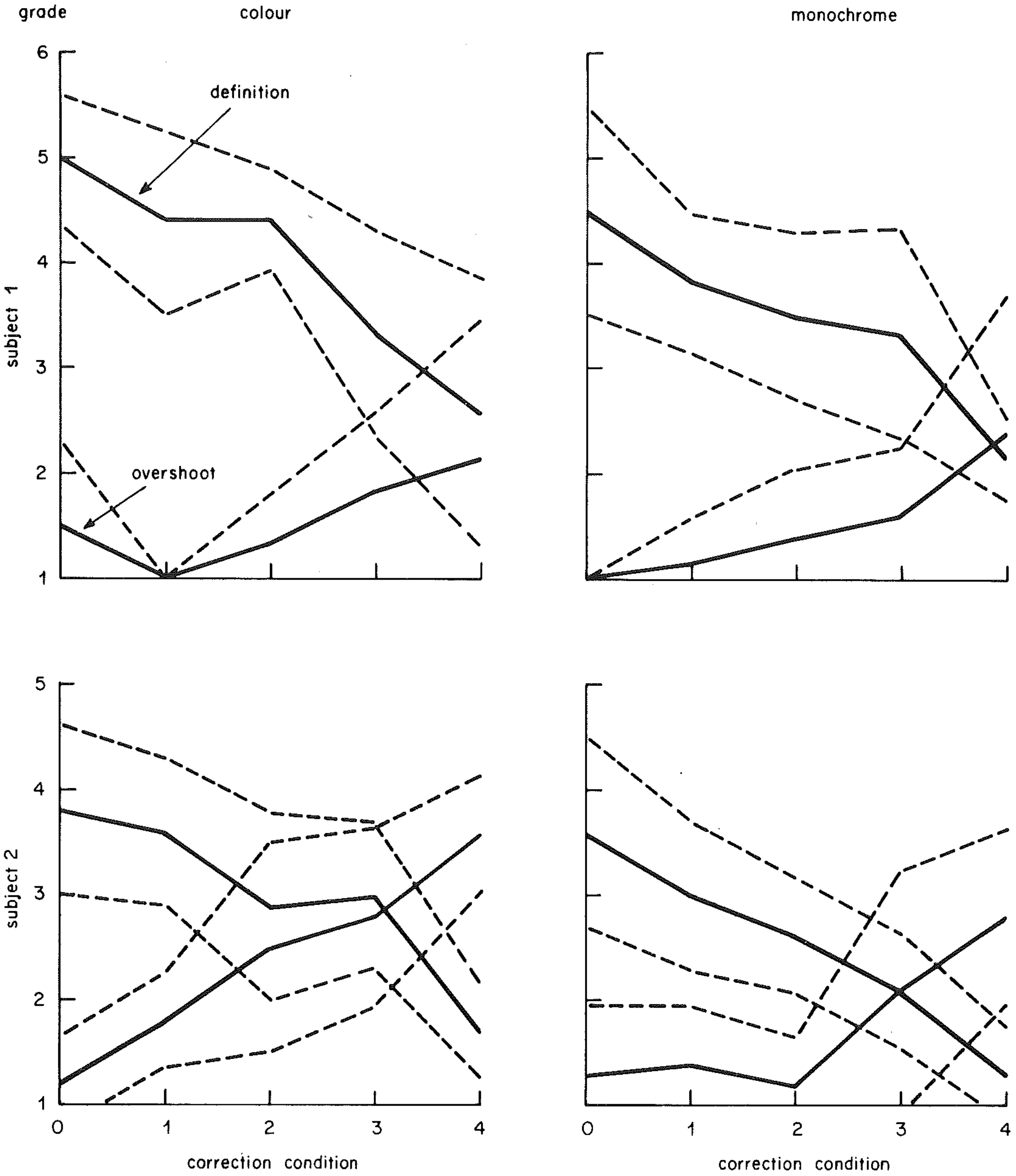


*Fig. 19 - Subjective grading of correction conditions for colour and monochrome pictures*

*TABLE 2*

*EBU Impairment Scale*

| Grade | Degree of Impairment |
|---|---|
| 1 | Imperceptible |
| 2 | Just perceptible |
| 3 | Definitely perceptible but not disturbing |
| 4 | Somewhat objectionable |
| 5 | Definitely objectionable |
| 6 | Unusable |

The results, averaged over observers, are shown in Fig. 19. The dashed lines indicate the standard deviation range. Table 1 shows that correction condition 4 corresponds to greater than 40 dB source loss at the half-field sampling frequency. As this is much greater than the loss of even the worst plumbicon tubes[7] this condition would be expected to overcorrect the pictures grossly. In fact the observers rated this condition as giving the best improvement in definition whilst giving the worst overshoot. The other conditions as might be expected, were rated in between this limit and that of giving no improvement. As can be seen there is a high inverse correlation between definition and overshoot and a high correlation between results for colour and monochrome pictures. When calculated for individual observers the self-correlation coefficient for colour and monochrome grading never fell below 0·9. It was thus concluded that this was sufficient justification for applying the results of monochrome picture tests to colour pictures.

### 3.3. The measurement of source aperture

The correction prediction method mentioned above requires a knowledge of the source aperture profile. It was decided to use a high-quality slide scanner as the source so that the source aperture, being the CRT spot profile, was, to some extent, under the control of the experimenter and would be relatively easy to measure. In fact there are many pitfalls in carrying out this measurement and it proved more difficult than anticipated.

There are three common ways of measuring a C.R.T. spot profile apart from direct measurement by microscope.

(a) The horizontal frequency method. In this method the spot scans a grating of known pitch and the light is collected by a photodetector whose output gives a measure of the response/frequency characteristic of the aperture at the grating frequency. By using different grating pitches the spectrum of the aperture can be built up and the Fourier transform gives the spot profile. If the grating is not sinusoidal care must be exercised in interpreting the photodetector output. The disadvantage of this method is that the spot must scan sufficiently slowly to eliminate errors caused by afterglow response in the C.R.T. phosphor. If, at normal scanning speeds, the afterglow correction is at all suspect, this method cannot be used.

(b) The knife-edge method. In this method the spot is scanned slowly across a knife-edge and the output of a photodetector collecting the transmitted light is proportional to the integral of the spot profile. Thus by differentiation the profile is obtained. The difficulty with this method is the maintenance of mechanical stability needed to produce repeatable results.

(c) The vertical frequency method. Here a spot scanning horizontally at normal television line frequencies is slowly scanned vertically across a horizontal grating and the photodetector output gives a spot measurement of the response/frequency characteristic as in (a). The difference between this method and that of (a) is simply the speed of scanning and the benefit of integration conferred by the horizontal scanning.

All these methods were tried and method (c) was found to be the most reliable. The vertical scanning was performed manually by operating the shift control on the scanner. The normal horizontal and vertical scans were left in operation and by strobing out a particular line of the photodetector output waveform the slow change in magnitude of the line signal could be observed as the shift was operated. The peak-to-peak magnitude of the signal gave a measure of the response at the grating frequency. Several grating frequencies were used (the gratings on Test Card C were found convenient) whose pitches were accurately measured. The results could be expressed in cycles/mm on the tube face knowing the magnification ratio of the optics associated with the scanner.

Fig. 20 shows the results obtained for various positions of the electronic focus control. The area of uncertainty accompanying response values near unity comes about because a square-wave grating was used. With such a grating a correction factor of $4/\pi$ must be applied to peak-to-peak measurements if the waveform is square but not if it is sinusoidal. The curves of Fig. 20 were subjected to best-fit analyses assuming Gaussian spectra and values for the standard deviation of the spot profile in $\mu$m were deduced shown in Table 3.

Referring to Section 2.3 the amount of correction required depends on the quantity $s$, the ratio of the spot size to the scanning pitch. The scanning pitch was found by measuring the number of field lines corresponding to a certain measured distance on a slide and converting this to a measurement on the tube face. The values of $s$ appropriate to the field scanning pitch thus deduced are shown in Table 3. This quantity can be used in the simultaneous

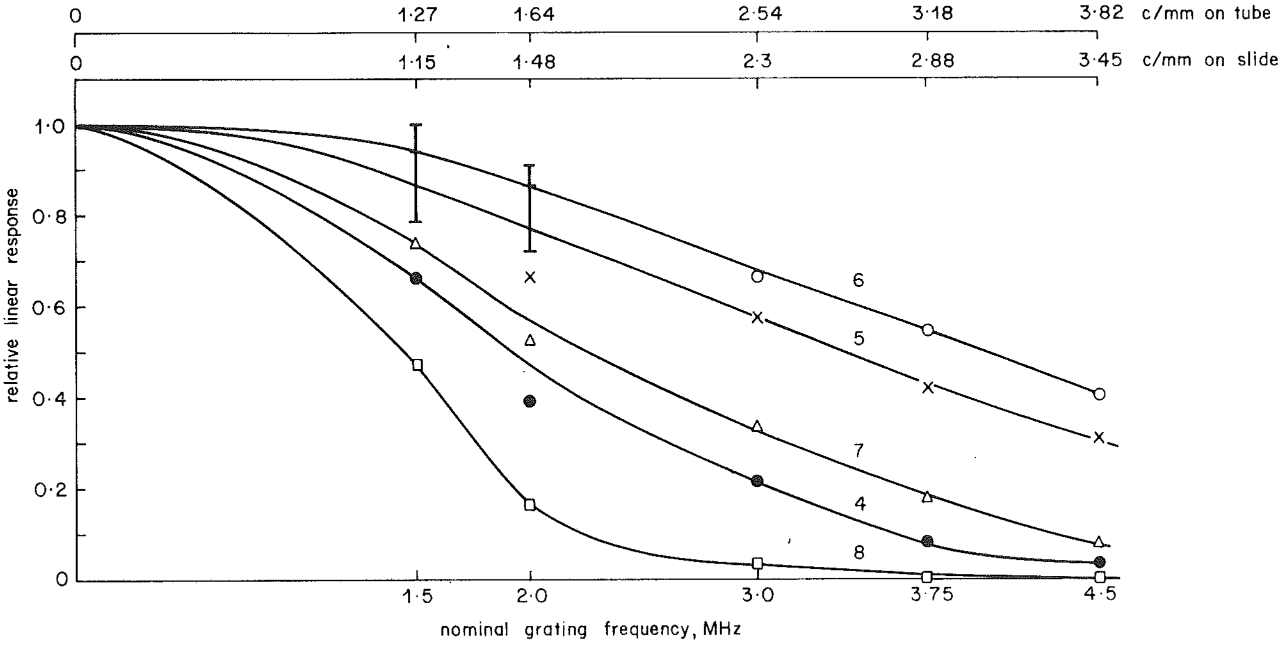


*Fig. 20 - Response/frequency characteristic of scanner for various focus conditions*

*TABLE 3*

*Source Characteristics*

| Source Condition | Spot standard deviation (µm) | $s_f$ | dB loss at half-sampling frequency | | Predicted correction coefficients | | | |
|---|---|---|---|---|---|---|---|---|
| | | | field | picture | $c_0$ | $c_1$ | $c_2$ | $c_3$ |
| 4 | 112 | ·66 | 18·6 | 74·4 | 1·436 | −·218 | | |
| | | | | | 1·693 | −·390 | ·043 | |
| | | | | | 1·859 | −·513 | ·092 | −·008 |
| 5 | 67 | ·40 | 6·85 | 27·4 | 1·163 | −·082 | | |
| | | | | | 1·224 | −·122 | ·010 | |
| | | | | | 1·255 | −·145 | ·019 | −·002 |
| 6 | 55 | ·325 | 4·5 | 18 | 1·104 | −·052 | | |
| | | | | | 1·138 | −·075 | ·006 | |
| | | | | | 1·156 | −·088 | ·011 | −·001 |
| 7 | 94 | ·555 | 13·2 | 52·8 | 1·308 | −·154 | | |
| | | | | | 1·458 | −·254 | ·025 | |
| | | | | | 1·544 | −·319 | ·051 | −·004 |
| 8 | 156 | ·925 | 36·5 | 146 | 1·849 | −·425 | | |
| | | | | | 2·603 | −·927 | ·126 | |
| | | | | | 3·254 | −1·415 | ·321 | −·033 |

equations to predict the correction coefficients because the transversal-filter delay element corresponds to the pitch of a field line. Also shown in Table 3 are the source losses at the half-field sampling and half-picture sampling frequencies. As the worst plumbicon tubes have a loss of 20 dB at the half-picture sampling frequency[7] it can be seen that only positions 5 and 6 correspond to probable source characteristics.

*TABLE 4*

*EBU Comparative Scale*

| Grade | Degree of Comparison |
|---|---|
| +3 | Much better than |
| +2 | Better than |
| +1 | Slightly better than |
| 0 | Same as |
| −1 | Slightly worse than |
| −2 | Worse than |
| −3 | Much worse than |

### 3.4. Subjective assessments of predicted correction

The coefficients necessary to correct the various source apertures according to the criterion of maximally-flat response are shown in Table 3. As the coefficients in the digital aperture corrector were quantised to 7-bit accuracy it would have been possible to apply 3rd-order correction for only conditions 4 and 8. Consequently it was decided to concentrate on differences between 1st and 2nd-order correction. Source conditions 4 and 5 were used and for each condition observers were asked to judge the difference between zero (no correction) and first-order, and between first and 2nd-order correction, according to the EBU Comparative Scale shown in Table 4. Three slides were used as source material; the first was a Test Card, the second a scene having random fine detail and a black outline (boat), and the third contained more ordered detail (Maida Vale Studio).

The results are shown in histogram form in Fig. 21. The hatched area represents the standard deviation and a single line indicates a unanimous score. For source condition 5 only Test Card was judged to be significantly improved by adding 2nd-order correction and only by a grade. For source condition 4, Test Card and Maida Vale were judged to be improved by a grade. The general comment was that insufficient correction had consistently been applied.

### 3.5. Subjective assessment of optimally corrected pictures

Following the results of the above test it was concluded that the criterion of maximally-flat response was not subjectively the most pleasing. A further test was therefore devised in which observers were first asked to indicate the degree of first-order correction they preferred. Two source conditions and two slides were used. This degree of correction was then held as a reference and the observers were asked to judge between this reference and various degrees of second-order correction using the EBU comparative scale. The amount of second-order correction was varied within a range which included the degree of first-order correction for low frequencies.

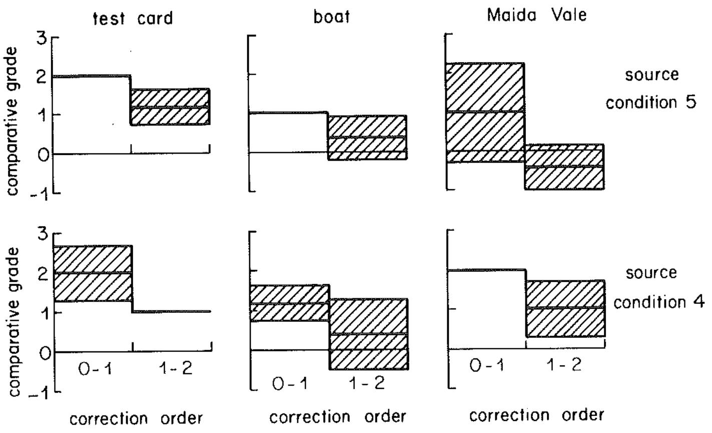


*Fig. 21 - Results of correction prediction subjective tests*
0 — 1 Comparison between no correction and first-order correction
1 — 2 Comparison between first order and second-order correction
Note: The sign of the comparative grade is positive if the last named condition is judged better than the first condition

The results of the first part of the test are shown in Table 5. The figures for losses at the half-field sampling frequency are those for the source, calculated from the coefficients according to the maximally-flat criterion. Comparing these figures with those of Table 3 calculated from actual source measurements it will be seen that there was always a tendency to over-correct although less so with the Test Card. Fig. 22 shows the response/frequency characteristic resulting from the coefficient values of Table 5 in conjunction with the source spectra calculated from source measurements.

This tendency to over-correct is not entirely understood but could have been caused in part by the display aperture which unfortunately was not measured. A display aperture spectrum with excessive attenuation at high spatial frequencies would require compensation in the signal fed to the display.

The results of the second part of the test are shown in Fig. 23 in which comparative grade (with respect to the standard) is plotted as a function of the two independent second-order coefficients, $c_1$ and $c_2$. It will be seen that for the great majority of conditions, second-order correction was judged inferior to first-order correction.

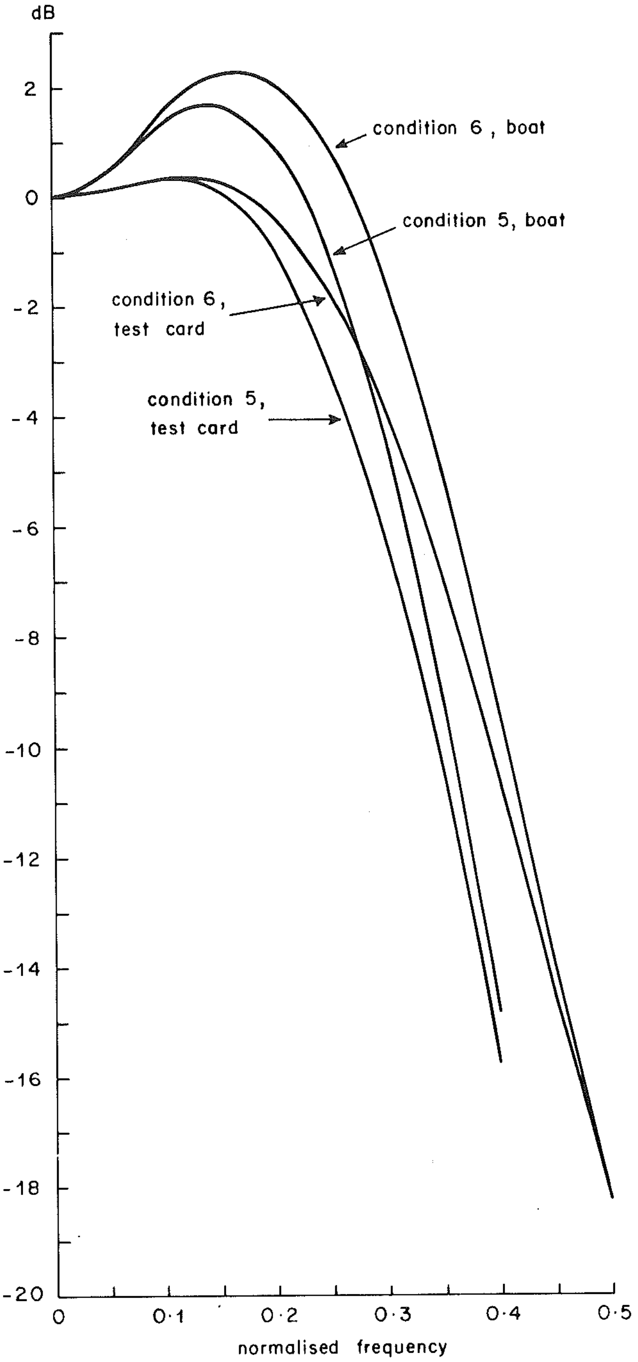


*Fig. 22 - Preferred overall response for two scenes and two source conditions*

*TABLE 5*

| Source Condition | Scene | Correction Coefficients | | dB loss at half-sampling frequency field | picture |
|---|---|---|---|---|---|
| 6 | Boat | 1·43 | —·215 | 18·4 | 73·6 |
| 5 | Boat | 1·46 | —·23 | 19·7 | 78·8 |
| 6 | Test Card | 1·19 | —·095 | 8·1 | 32·4 |
| 5 | Test Card | 1·26 | —·13 | 11·1 | 44·4 |

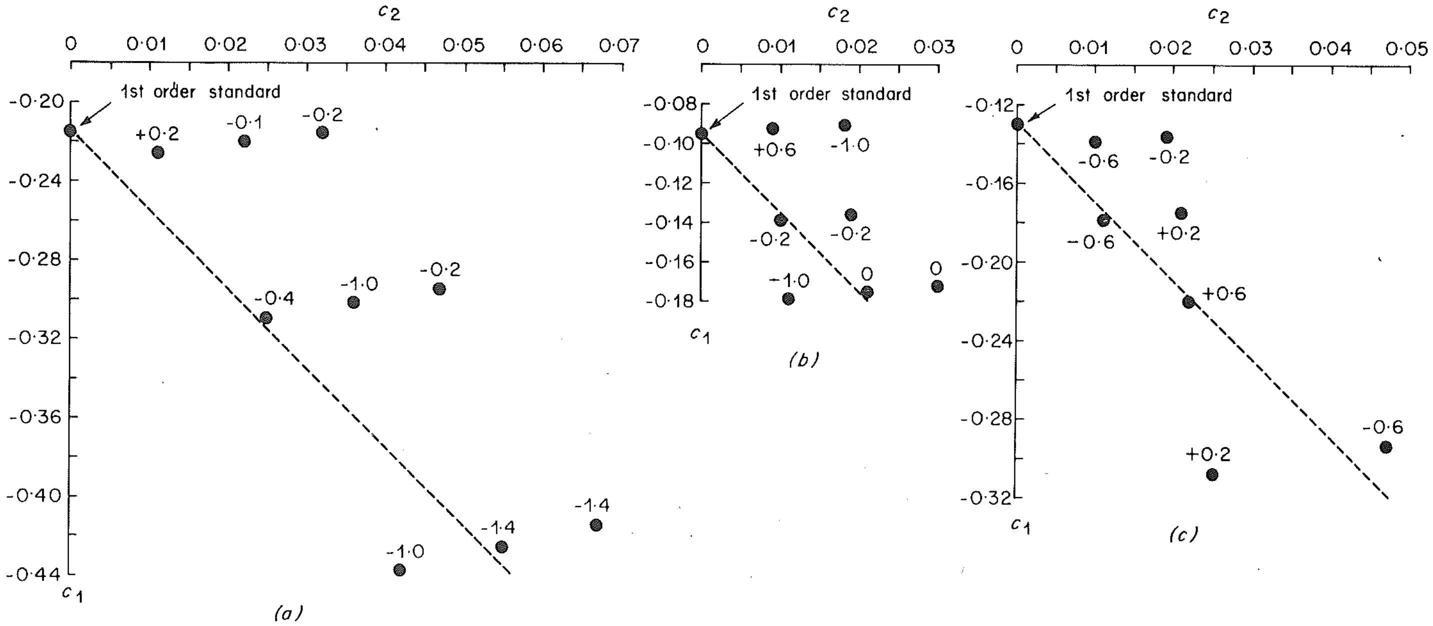


*Fig. 23 - Subjective assessment of second-order correction by comparison with first-order preferred correction. (a), (b) and (c) represent different source conditions and picture material. Comparative grade is plotted on a three-dimensional graph as a function of the two independent second-order variables. The plotted points give the values of $c_1$ and $c_2$ appropriate to the second-order correction conditions and the numbers beside the points are values of the subjective grading of the conditions. Note the grading is positive if the condition is judged better than the standard. The dotted line indicates the relationship between $c_1$ and $c_2$ when the response/frequency characteristic of the overall system has the same behaviour near the frequency origin as the standard condition*

## 4. Conclusions

It would appear from both theoretical and practical considerations that there is little to be gained from increasing the complexity of aperture correction beyond first-order. A relatively crude analysis of perfect correction reveals that the correction coefficient values decrease as the inverse square of the distance from the wanted signal. An analysis of practical correction indicates that the rate of decrease is even faster. Consequently the higher-order terms rapidly become insignificant. This law of diminishing returns is confirmed by the theoretical time-domain response to various edges and especially in the case of interlaced-sampling where higher order correction may actually impair the response by introducing ringing before and after edges. Subjective tests indicate that there is a slight preference for second-order correction over first-order if the correction is of such magnitude that the frequency response is maximally-flat at zero frequency, and does not overshoot. However, first-order over-correction is preferred to either of these correction conditions.

These subjective results are based on tests using a corrector operating on the signals related to a single field. If, in an interlaced system, it were possible to operate on the signals of both fields there might be a slight gain in introducing second-order correction but this would only apply for stationary scenes. Even for these scenes, using such refined correction, there would probably be difficulty with 'twittering' on horizontal edges. This is purely a defect of conventional display devices which rely on the fact that in an interlaced system the picture content does not exploit the full spatial bandwidth of the two combined fields.

For moving scenes, at least a picture store would be required in order to interpolate points in the interlaced field and the present cost of this process would probably be deemed excessive in relation to the benefit gained.

## 5. References

1. GIBSON, W.G. *and* SCHROEDER, A.C. 1960. A vertical aperture equalizer for television. *J. Soc. Motion Pict. Telev. Engrs,* 1960, **69**, 6, pp. 395 – 401.

2. WALKER, R. 1971. Digital line-store standards conversion: a feasibility study. BBC Research Department Report No. 1971/44.

3. Digital line-store standards conversion: determination of the optimum interpolation aperture function. BBC Research Department Report in course of preparation.

4. MERTZ, P. *and* GRAY, F. 1934. A theory of scanning and its relation to the characteristics of the transmitted signal in telephotography and television. *Bell Syst. Tech. J.*, 1934, **13**, 3.

5. DEVEREUX, V.G. 1966. The choice of interpolation apertures for line-store standards converters. BBC

Research Department Report No. T-177, Serial No. 1966/61.

6. HOWORTH, D. 1963. A method of calibration of an experimental vertical aperture corrector. BBC Research Department Report No. T-097/2, Serial No. 1963/16.

7. BBC specification for photo-conductive, lead-oxide target, camera tubes. Specification TV/153.

## Appendix 1

## The Derivation of Correction Coefficients for Perfect Source Correction

Following the method outlined in Section 2.3 it is assumed that the correction-aperture spectrum can be expressed as an even power series in the spatial frequency variable, that is

$$F(\nu) = N^{-1}[1 + (\nu/\nu_2)^2 + (\nu/\nu_4)^4 + \ldots\ldots], \qquad |\nu| < N/2$$

$$= 0 \text{ otherwise}$$

where $N$ is the spatial scanning frequency. The Fourier transform can be written

$$f_0(y) + f_2(y) + \ldots\ldots = \sum_{k=0}^{\infty} f_{2k}(y)$$

where $f_{2k}$ results from the term in $\nu^{2k}$. In general

$$f_{2k}(y) = N^{-1}\nu_{2k}^{-2k}\, I_{2k}(y) \qquad \text{where}$$

$$I_{2k}(y) = \int_{-N/2}^{N/2} \nu^{2k} \exp 2\pi \mathrm{j}\nu y \;\, \mathrm{d}\nu$$

Now the identity

$$\int_{-1}^{+1} \nu^k \exp - \mathrm{j}y\nu \;\mathrm{d}\nu = \mathrm{j}^k \frac{\mathrm{d}^k}{\mathrm{d}y^k}\left(\frac{2\sin y}{y}\right)$$

is derivable from a standard general form. Thus

$$I_{2k}(y) = \int_{-N/2}^{N/2} \nu^{2k} \exp 2\pi \mathrm{j}\nu y \;\mathrm{d}\nu = N\frac{\mathrm{j}^{2k}}{(-2\pi)^{2k}} \frac{\mathrm{d}^{2k}}{\mathrm{d}y^{2k}}\left(\frac{\sin N\pi y}{N\pi y}\right)$$

$$= N(-1)^k\left(\frac{N}{2}\right)^{2k} \frac{\mathrm{d}^{2k}}{\mathrm{d}z^{2k}}\left(\frac{\sin z}{z}\right)$$

$$= N\left(\frac{N}{2}\right)^{2k}\left[\frac{\sin z}{z} + \frac{2k\cos z}{z^2} - 2k(2k-1)\frac{\sin z}{z^3} - \ldots + (-1)^k(2k)!\,\frac{\sin z}{z^{2k+1}}\right]$$

where $z = N\pi y$

The coefficients are points on the correction aperture spaced at intervals of $1/N$, that is at the points $r/N$ where $r$ is an integer.

Thus
$$c_r = \sum_{k=0}^{\infty} f_{2k}(r/N) = \sum_{k=0}^{\infty} N^{-1}\nu_{2k}^{-2k}\, I_{2k}\,(r/N)$$

and

$$I_{2k}(r/N) = N\left(\frac{N}{2}\right)^{2k}\left[\frac{\sin\pi r}{\pi r} + \frac{2k\cos\pi r}{(\pi r)^2} - 2k(2k-1)\frac{\sin\pi r}{(\pi r)^3} - \ldots + (-1)^k(2k)!\frac{\sin\pi r}{(\pi r)^{2k+1}}\right]$$

$$= N\left(\frac{N}{2}\right)^{2k}(-1)^r\left[\frac{2k}{(\pi r)^2} - \frac{2k(2k-1)(2k-2)}{(\pi r)^4} + \ldots\ldots + \frac{(-1)^{k-1}(2k)!}{(\pi r)^{2k}}\right],\ r \neq 0,\ k \neq 0$$

For $r = 0$

$$I_{2k}(0) = N(-1)^k\left(\frac{N}{2}\right)^{2k}\lim_{z\to 0}\frac{d^{2k}}{dz^{2k}}\cdot\left(1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \ldots\ldots\right)$$

Thus $I_0(0) = N$

$$I_2(0) = N(-1)\left(\frac{N}{2}\right)^2\lim_{z\to 0}\left(\frac{-2}{3!} + \frac{4.3\,z^2}{5!} - \ldots\right) = N\left(\frac{N}{2}\right)^2\frac{2}{3!}$$

$$I_4(0) = N(-1)^2\left(\frac{N}{2}\right)^4\lim_{z\to 0}\left(\frac{4!}{5!} - \frac{6.5.4.3.z^2}{7!} + \ldots\right) = N\left(\frac{N}{2}\right)^4\frac{4!}{5!}$$

$$I_{2k}(0) = N\left(\frac{N}{2}\right)^{2k}\frac{1}{2k+1}$$

Therefore

$$c_0 = \sum_{k=0}^{\infty}\frac{1}{2k+1}\left(\frac{N}{2\nu_{2k}}\right)^{2k}$$

$$c_r = \sum_{k=1}^{\infty}(-1)^r\left(\frac{N}{2\nu_{2k}}\right)^{2k}\left[\frac{2k}{(\pi r)^2} - \frac{2k(2k-1)(2k-2)}{(\pi r)^4} + \ldots\ldots\right]$$

$$= (-1)^r\sum_{k=1}^{\infty}\left[\left(\frac{N}{2\nu_{2k}}\right)^{2k}\sum_{q=1}^{k}\frac{(2k)!}{(2k+1-2q)!}\,\frac{(-1)^{q-1}}{(\pi r)^{2q}}\right]$$

For the Gaussian aperture of standard deviation $\sigma$

$$F(\nu) = N^{-1}\exp 2\pi^2\sigma^2\nu^2 \qquad |\nu| < N/2$$

$$= 0 \qquad \text{otherwise.}$$

But

$$\exp 2\pi^2\sigma^2\nu^2 = 1 + 2\pi^2\sigma^2\nu^2 + (2\pi^2\sigma^2\nu^2)^2/2! + \ldots.$$

so that

$$\nu_{2k}^{-2k} = (2\pi^2\sigma^2)^k/k!$$

Thus

$$c_0 = \sum_{k=0}^{\infty} \frac{1}{2k+1}\,\frac{(2\pi^2\sigma^2)^k}{k!}\left(\frac{N}{2}\right)^{2k} = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k+1)k!} \tag{1a}$$

and

$$c_r = (-1)^r \sum_{k=1}^{\infty} \left[\frac{z^{2k}}{k!} \sum_{q=1}^{k} \frac{(2k)!}{(2k+1-2q)!}\,\frac{(-1)^{q-1}}{(\pi r)^{2q}}\right] \tag{1b}$$

where $z = \pi\sigma N/\sqrt{2} = \pi\ s/\sqrt{2}$

For the rectangular aperture of width $w/N$

$$\left.\begin{aligned} F(\nu) &= N^{-1}\left(\frac{\pi w\nu/N}{\sin \pi w\nu/N}\right), \qquad && |\nu| < N/2 \\ &= 0 && \text{otherwise} \end{aligned}\right\} \quad w < 2$$

Now it can be shown that the expansion of $x/\sin x$ has the form

$$x/\sin x = 1 + A_2 x^2 + A_4 x^4 + \ldots\ldots\ldots$$

where

$$A_2 = 1/3!$$

$$A_4 = 1/(3!)^2 - 1/5!$$

$$A_6 = 1/(3!)^3 - 2/3!5! + 1/7! \text{ etc.}$$

The ratio of successive coefficients tends to $\pi^2$.

Thus

$$F(\nu) = N^{-1}\left[1 + A_2\,(\pi w\nu/N)^2 + A_4\,(\pi w\nu/N)^4 + \ldots\ldots\right]$$

and

$$\nu^{2k}_{2k} = (N/\pi w)^{2k}/A_{2k}$$

Thus

$$c_0 = \sum_{k=0}^{\infty} \frac{A_{2k} z^{2k}}{2k+1} \tag{2a}$$

and

$$c_r = (-1)^r \sum_{k=1}^{\infty} \left[A_{2k} z^{2k} \sum_{q=1}^{k} \frac{(2k)!}{(2k+1-2q)!}\,\frac{(-1)^{q-1}}{(\pi r)^{2q}}\right] \tag{2b}$$

where $z = \pi w/2$.

The values of $z$ can be related to $D$, the dB of source loss at the frequency $N/2$ by the equation

$$D = 20 \log_{10} [F(N/2)/F(0)]$$

For the Gaussian source

$$
\begin{aligned}
D &= 20\log_{10}\ \exp 2\pi^2\sigma^2N^2/4 \\
&= 8{\cdot}68\ldots.\quad \pi^2s^2/2 \\
&= 8{\cdot}68\ldots.\quad z^2 \qquad (3)
\end{aligned}
$$

For the rectangular source

$$
\begin{aligned}
D &= 20\log_{10}\ [(\pi w/2)/\sin(\pi w/2)] \\
&= 20\log_{10}\ (z/\sin z) \qquad (4)
\end{aligned}
$$

By eliminating $z$ in Equations (1) and (3) or (2) and (4) the coefficient values can be deduced as functions of source loss.

## Appendix 2

## The Derivation of Coefficient Values for Finite-order Correction

For finite-order correction on a maximally-flat basis ascending powers of $[Y_y(n)]^{-1}$ and $H(\theta)$ are equated. Assuming a Gaussian source we have

$$[Y_y(n)]^{-1} = \exp 2\pi^2\sigma^2n^2$$

For $l$th order correction

$$H(\theta) = c_0 + 2\sum_{r=1}^{l} c_r \cos r\theta$$

where $\theta = 2\pi n/N$

Thus equating power series

$$
\begin{aligned}
&1 + 2\pi^2\sigma^2n^2 + (2\pi^2\sigma^2n^2)^2/2!\ldots\ldots \\
&= c_0 + 2c_1(1-\theta^2/2! + \theta^4/4! + \ldots) + 2c_2(1 - 2^2\theta^2/2! + 2^4\theta^4/4! + \ldots) \\
&\quad + \ldots\ldots\ldots
\end{aligned}
$$

which leads to the simultaneous equations

$$
\begin{aligned}
&c_0/2 + c_1 + c_2 + \ldots\ldots + c_l = \tfrac{1}{2} \\
&\quad c_1 + 2^2c_2 + \ldots. + l^2c_l = -\tfrac{1}{2}\,(2!/1!)2^{-1}(\sigma N)^2 \\
&\quad \ldots\ldots\ldots. \\
&\quad c_1 + 2^{2q}c_2 + \ldots + l^{2q}c_l = \tfrac{1}{2}(-1)^q[(2q)!/q!]\ 2^{-q}(\sigma N)^{2q} \\
&\quad \text{etc.}
\end{aligned}
$$

Thus we have an infinite number of equations with only $l + 1$ unknowns. On a maximally-flat basis the first $l + 1$ equations are satisfied giving values of the coefficients $c_0 \ldots c_l$ as a function of the parameter $\sigma N$ (= $s$, the normalised standard deviation).

SMW/JUC

Printed by BBC Research Department, Kingswood Warren, Tadworth, Surrey